An in silico analysis: Deciphering the role of differential expressed genes in tuberculosis in response to bedaquiline antibiotic drug

K. Santhiya*, V. Anusuya, P. Madhan, B. Nivetha
Department of Biotechnology, Faculty of Engineering, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India.

ARTICLE INFO

Article history:
Received on: January 11, 2022
Accepted on: March 30, 2022
Available online: ***

Key words: Bedaquiline, Functional enrichment, Hub genes, Interaction, Tuberculosis

ABSTRACT

Tuberculosis (TB) caused by the bacteria Mycobacterium TB (MTB) is one of the most devastating respiratory infections. The drug Bedaquiline, which is used to treat tuberculosis, was developed to improve treatment outcomes while reducing the toxicity associated with injectable medicines. Clinical prescriptions for the medication have increased in recent years, but its therapeutic efficacy has decreased. Drug resistance arises when MTB become resistant to the medications used to treat TB, and these treatments no longer suppress the gene activity. In the present study, an integrated bioinformatics analysis was performed to explore potential crucial genes, key pathways and interaction of miRNA with the transcription factor associated with the differentially expressed gene in TB in response to a multidrug resistant TB drug named bedaquiline. We have curated the dataset GSE43764 and carried out the functional and pathway enrichment analysis using various bioinformatics tool. The genes encoding dnaA, dnaB, dnaE1, atpE, atpF, and atpB were identified as hub genes. MiR-574-3p, MiR-4787-5p, MiR-601, and MiR-1234 have been found as possible target microRNA. The potential transcription factors pertaining to the Bedaquiline and TB were also found. The results indicate that the discovered key differentially expressed gene has a promising future and can be employed as biomarkers.

1. INTRODUCTION

Tuberculosis (TB) is one of the world’s most lethal illnesses. Every day, about 4000 people die as a result of TB. The enormous disruptions in health systems created by the COVID-19 pandemic are likely to escalate the dreadful toll by 1 million TB fatalities per year over the next 4 years [1]. TB is a dangerous infectious illness that primarily affects the lungs and is caused by Mycobacterium TB (MTB). TB affects one-quarter of the world’s population, with 10.4 million new cases reported each year, despite the widespread use of a live attenuated vaccine and a variety of treatments. As an aerosol droplet, MTB penetrates the alveolar airways of exposed persons, where it is thought to establish its first contact with resident macrophages [2]. The COVID pandemic, the emergence of multidrug-resistant (MDR) TB, and communities disproportionately restricted access to treatment all impact the worldwide burden of TB.

Both extensively drug-resistant TB (isoniazid, rifampin, and fluoroquinolone) and injectable drug (capreomycin or kanamycin) and complicated forms of MDR TB were shown to be ineffective in controlling threats of TB. Bedaquiline, a novel diarylquinoline, was added to the WHO-recommended all-oral regimen to replace injectable treatments for MDR-TB patients. The WHO consolidated guidelines on drug-resistant TB treatment [3]. Bedaquiline has been reported to be a mycobacterial adenosine triphosphate (ATP) synthase inhibitor; it binds and disrupts the FO subunit interface, resulting in an inefficient proton cycle, which is lethal to Mycobacterium [4]. When compared to comparable eukaryotic enzymes, the mycobacterial ATP synthase enzyme is highly selective (Selectivity Index >20, 000) [5]. As a result, many people infected with TB remain misdiagnosed or are treated without test confirmation. Postmortem investigations demonstrated that TB is a common undetected cause of death, highlighting the need for better diagnostic tools. Microarray-based RNA expression analysis has evolved into an essential tool for studying disease biology. Many illnesses, such as cancer, infectious disorders, arthritis, HIV, and TB, are linked to distinct transcriptional patterns in blood or tissue [6].

It uses gene expression profiling to characterize complicated cellular responses, and new molecular pathways have been discovered under various circumstances. Because of the abundance of publicly available data, it is possible to pool gene expression datasets and boost sensitivity by increasing the number of data points. This method has already been utilized to discover gene signatures and pathways that are coregulated in various disorders. Similarly, a meta-analysis of gene expression has helped to discover new genes and pathways that are deregulated inactive TB [7]. We conducted gene interaction network studies to investigate the role of differentially expressed genes (DEGs) at a molecular level in the present study, which utilized gene expression data during exposure to antituberculosis medications, bedaquiline. The
gene expression omnibus GEO2R tool and R programming were used to perform differential expression analysis. The functional enrichment analysis was performed for the DEGs to find the hub genes and enhanced pathways, which will help researchers, better comprehend the MTB system-wide roles following bedaquiline exposure. The study’s goal was to evaluate and identify possible biomarkers that might aid in the diagnosis of TB from a worldwide TB pandemic era.

2. METHODOLOGY

2.1. Retrieval of Datasets

The gene dataset was retrieved using GEO database using keywords: TB and bedaquiline and *Homosapiens* (https://www.ncbi.nlm.nih.gov/gds/?term=tb+and+bedaquiline) [8]. The NCBI-Gene expression omnibus database (NCBI-GEO) is facilitated with the analyses by furnishing several microarray datasets. GSE43749 datasets were chosen for the study as per the inclusion criteria framed.

2.2. Datasets Processing

GEO2R is a web-based interactive tool to compare two different data sets (Normal and Patient group). It is effectively used to pick out DEGs from the given dataset. It will analyze by assigning the group types. For the dataset GSE43749, 251 DEGs were obtained. The results obtained consist of gene ID, gene symbol, gene title, P-value, adj. P-value, and log FC value of DEG (https://www.ncbi.nlm.nih.gov/geo/geo2r/).

2.3. Identification of DEGs

R programming was performed to obtain upregulated and downregulated genes for the given dataset. The cutoff value for screening the DEGs was considered by fixing the log fold change ranges between –0.5 and 0.5 and P ≤ 0.05. The result was obtained using an R programming code consisting of 38 upregulated genes and 37 downregulated genes.

2.4. Functional Annotation and Pathway Enrichment Analysis

The enrichment analysis was performed to explore the biological implications of DEGs. The functional enrichment of DEG in cellular functionality, biological processes, and molecular functions was identified using the PANTHER database. It is a bioinformatics tool to study the biological information for a large set of genes, and functional annotation can be done using an integrated gene ontology database [9]. Pathway enrichment analysis was introduced from genome-scale trials to better understand its features, and it finds biological pathways that are enriched in a gene list provided as input. EnrichR is a comprehensive resource for curated gene sets and a search engine that accumulates biological information for subsequent biological discoveries of DEGs, which is used in this study [10].

2.5. Comprehensive Analysis of Protein-protein Interactions (PPI) Networks and Modules

After removing redundant genes, the collection of genes generated from the preceding stages was pooled together. PPIs are crucial for understanding how biological systems are regulated. PPI has been found in recent research to identify critical hub genes that cause illnesses and also act as therapeutic targets for the precise treatment of a drug over the corresponding disease. The hub genes mainly play a key role in the potential pathway analysis. The STRING database was used to perform PPI among uniquely discovered DEGs on the whole human genome [11]. The DEGs were mapped to the STRING database with a confidence score of 0.6, and the results were visualized using Cytoscape. The molecular complex detection (MCODE) plugin was used to filter the key network modules with degree cutoff = 10, node score cutoff = 0.2, k core = 2, and max depth = 100 [12].

2.6. Construction of Target Genes miRNA – Regulatory Network

The potential microRNAs were evaluated for correlations with the candidate genes to uncover possible microRNA-gene correlations. The increased target-gene expression may be linked to increased microRNA expression and vice versa. To create an in-depth relationship, a link between the microRNA profile and gene regulatory networks of both mouse and human by collecting and integrating the documented regulatory interaction was found using the EnrichR tool. The Regenetwork tool was used to understand the details of the transcription factor associated with the human miRNA with the target score and the gene symbol [13]. Finally, the Chea3 tool was used to get the clustering association of the transcription factors involved with the respective miRNA [14].

3. RESULTS AND DISCUSSION

3.1. Identification of DEGs

To find DEGs, the gene expression dataset related to TB was found by searching in the GEO database using the keywords: TB, bedaquiline, and *Homosapiens*. The dataset which satisfies our inclusion criteria was GSE43749; it was taken for the subsequent processing steps. The GEO datasets are compared using the GEO2R tool and the DEGs are identified from the whole gene sets [Figure 1]. According to the [log2FC] ±0.5 and P < 0.05, 38 are upregulated genes and 37 are downregulated genes that were observed out of 251 DEG’s by performing R programming.

3.2. Functional Annotation

We used PANTHER to enrich the procured set of DEGs to get the potential GO categories by classifying them by molecular function, cellular function, biological process, and protein class to determine the importance of the detected DEGs. The identified DEGs were primarily associated with molecular functions divided into the binding, catalytic activity, and transporter activity. It also indicates that both upregulated and downregulated genes have a high amount of catalytic activity, as per Figure 2 and 3. A biological process component of GSE43749 dominantly belongs to the cellular process, metabolic process, localization, and biological response. It was predicted that both cellular and metabolic processes are involved in the essential life processes. The cell envelope of MTB is complex and it is mainly composed of peptidoglycans, mycolic acids, lipids, and carbohydrates. Thus, it is conceivable that the cellular processes of MTB genes belonging to this class could play a key role in mycobacterial intracellular functions and cellular anatomical entry, which can be predominantly used as potential drug targets [15].

The protein complex possesses a structural molecular activity in binding the target drug with the genes. The protein classes of GSE43749 include chemokine, cytokine, hydrolase, ribosomal protein, and RNA-binding protein that were enriched along with the signaling and cell adhesion molecule. The curated DEGs were enriched in the gene-specific transcriptional regulator, metabolite inter-conversion enzyme, nucleic acid metabolism protein, and transporters. The genes required for optimal growth of the MTB are identified to be involved in several central metabolic pathways.
3.3. Pathway Enrichment Analysis

The pathway analysis of DEGs was identified using the EnrichR tool. The result served that Fc epsilon RI signaling pathway enriched for the dataset taken followed by other pathways in Table 1. Fc epsilon RI-mediated signaling pathways in mast cells are initiated by the interaction of antigen (Ag) with immunoglobulin (Ig)E bound to the extracellular domain of the alpha chain of Fc epsilon RI.

Fc epsilon RI-mediated signaling pathways are initiated by the interaction of Ag with antibody (IgE) bound to the extracellular domain of the alpha chain of Fc epsilon RI. The activation pathways are regulated positively and negatively by the interactions of many signaling molecules. Mast cells that are activated thus release granules that contain biogenic amines and proteoglycans. The activation of phospholipase A2 causes the release of membrane lipids, which develops lipid mediators such as leukotrienes (LTC4, LTD4, and LTE4) and prostaglandins (especially PDG2). There is also a secretion of cytokines, the most important of which are TNF-alpha, IL-4, and IL-5. These mediators and cytokines contribute to inflammatory responses [16].

3.4. Identification and Validation of Hub Genes

We used Cytoscape to execute a PPI network analysis based on the STRING interactome database to investigate the interactive interactions between the DEGs that are represented in Figure 4. The DEGs were formed as two major clusters during Cytoscape MCODE plug-in analysis. One cluster encodes the hub gene dnaA, dnaB, and dnaE1 [Figure 5a], whereas another cluster has atpE, atpF, and atpB [Figure 5b].

The dnaA protein interacts with repeated and non-palindromic DNA boxes located inside the oriC region to initiate bacterial chromosomal replication. The associations of dnaA protein with MTB DNA may provide new light on the protein’s function and aid in understanding the control of mycobacterial chromosomal replication start [17]. In Mycobacteria, dnaB is the replicative helicase, which is essential in the replication and growth process. dnaB is also one of the most often seen intein-containing proteins in bacteria [18]. dnaE1 and dnaE2, two DNA polymerases involved in replication in MTB, appear functionally identical. dnaE2 is required for damage-induced base-substitution mutagenesis in MTB. MTB becomes hypersensitive to DNA damage when dnaE2 activity is lost, induced mutagenesis is eliminated, virulence is reduced, and drug resistance is reduced in vivo.

Along with dnaE2 and recA, dnaE1 provides critical and high-fidelity replicative polymerase activity and is expressed in response to DNA damage [19]. The genes atpE, atpF, and atpB belonged to ATP synthase. ATP synthase is a crucial enzyme in practically all living cells’ energy metabolism. In certain species, ATP synthase has distinct traits that might adapt to the habitats they encounter. Pathogenic bacteria may confront unique obstacles regarding ATP generation since they must deal with low oxygen tensions and nutritional scarcity [20].

3.5. miRNA Prediction

MicroRNAs are indigenous RNAs with roughly 22 bases that target mRNAs for cleavage or post-translational modifications and serve as an essential regulatory function in plants and animals. The miRNA identification is made using the Enrichr tool, and the results are shown to consist of human and mouse miRNA. The four human miRNA (hsa-miR-574-3p, hsa-miR-4787-5p, hsa-miR-601, and hsa-miR-1234) was preceded further for our study in Table 2.

The hsa-miR-574-3p is found in the human chromosome 4 intron regions. A recent study has discovered that changes in miR-574-5p expression are linked to various disorders, notably myocardial infarcted, colorectal, and lung cancer. TLR9 signaling may be enhanced by miR-574-5p in lung cancer, promoting tumor growth. MiR-574 expression was elevated in colorectal cancer and myocardial infarcted leading to enhanced proliferation and invasion, whereas in this study, the expression of miR-574 was downregulated in response with TB [21]. MicroRNAs and cardiac sarcoplasmic reticulum calcium ATPase-2 in human myocardial infarction: Expression and

Figure 1: Volcano plot for control and patient samples of GSE43749.

Figure 2: Functional annotation of upregulated gene set.

Figure 3: Functional annotation of downregulated gene set.
bioinformatics analysis. The study by Budak et al., 2018 determined that miRNA-4787 showed a similar expression pattern in both acute and chronic groups in interaction with CD8 cells, whereas another study by Pu et al., 2016 reveals that miR-4787 was consistently expressed in tissues and plasma of patients with various stages of lung carcinoma [22,23].

The hsa-miR-4787-5p and hsa-miR-601 can be used to target gene-related cytokines. The miRNAs are involved in several physiological and pathological processes. It is a positive modulator of TLR signaling and is activated when murine macrophages are infected with *Mycobacteria*.

### 3.6. Comprehensive Analysis of miRNA with Transcription Factor

The Regnetwork was used to find possible transcription factors involved in miRNA (hsa-miR-574-3p, hsa-miR-4787-5p, hsa-miR-601, and hsa-miR-1234) that were then evaluated to determine TF coexpression networks using the chea3 tool.

<table>
<thead>
<tr>
<th>Serial number</th>
<th>Name</th>
<th>P</th>
<th>Adjusted P value</th>
<th>OR</th>
<th>Combined score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fc epsilon RI signalling pathway</td>
<td>0.08479</td>
<td>0.1954</td>
<td>11.88</td>
<td>29.33</td>
</tr>
<tr>
<td>2</td>
<td>Fc gamma R-mediated phagocytosis</td>
<td>0.1119</td>
<td>0.1954</td>
<td>8.84</td>
<td>19.36</td>
</tr>
<tr>
<td>3</td>
<td>Th1 and Th2 cell differentiation</td>
<td>0.1130</td>
<td>0.1954</td>
<td>8.74</td>
<td>19.05</td>
</tr>
<tr>
<td>4</td>
<td>NF-kappa B signaling pathway</td>
<td>0.1165</td>
<td>0.1954</td>
<td>8.46</td>
<td>18.19</td>
</tr>
<tr>
<td>5</td>
<td>T-cell receptor signaling pathway</td>
<td>0.1234</td>
<td>0.1954</td>
<td>7.95</td>
<td>16.63</td>
</tr>
<tr>
<td>6</td>
<td>Th17 cell differentiation</td>
<td>0.1303</td>
<td>0.1954</td>
<td>7.50</td>
<td>15.28</td>
</tr>
<tr>
<td>7</td>
<td>Natural killer cell-mediated cytotoxicity</td>
<td>0.1572</td>
<td>0.2021</td>
<td>6.11</td>
<td>11.30</td>
</tr>
<tr>
<td>8</td>
<td>Rap1 signaling pathway</td>
<td>0.2361</td>
<td>0.2618</td>
<td>3.86</td>
<td>5.57</td>
</tr>
<tr>
<td>9</td>
<td>Ras signaling pathway</td>
<td>0.2618</td>
<td>0.2618</td>
<td>3.42</td>
<td>4.58</td>
</tr>
</tbody>
</table>

OR: Odds ratio.

### Table 2: Details of human microRNA.

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Target score</th>
<th>Target site</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-574-3p</td>
<td>92</td>
<td>Retinoid and receptor alpha</td>
<td>The receptors function as transcription factors by to specific sequences in the promoters of target genes</td>
</tr>
<tr>
<td>hsa-miR-4787-5p</td>
<td>89</td>
<td>Cell migration inducing hyaluronidase2</td>
<td>It binds hyaluronic acid and catalyzes the depolymerization activity</td>
</tr>
<tr>
<td>hsa-miR-601</td>
<td>96</td>
<td>BCL2</td>
<td>MTB involves preventing apoptosis induced by up-regulation of Bcl2</td>
</tr>
<tr>
<td>hsa-miR-1234</td>
<td>87</td>
<td>Innate immunity activator</td>
<td>The development of a tuberculosis drug mainly targets the induction of adaptive immune responses</td>
</tr>
</tbody>
</table>

MTB: *Mycobacterium tuberculosis*, miRNA: Micro RNA.

The network analysis for TF with hsa-miR-574-3p was found to have 10 top TF, in which JUND, CAMTA2, and MEF2D were identified to have more interconnections with other TFs shown in Figure 6a. JUND belong to the JUN family of proteins. JUN and other TF can influence immunological activity in the body, and their gene function should be primarily extracellular and protein binding. The expression level JUN will be higher for people affected by SLE peripheral blood mononuclear cells infection [24].

The miRNA, hsa-miR-4787-5p, showed the higher expression of CEBP and JUND in Figure 6b. An integrated study by Delgado et al., 2019 revealed that TB patients tend to have disease severity with the presence of the CEBP gene in connection with IL6 [25]. The network analysis of a set of TFs for the-miR-601 had shown the central
clustering with Early Growth Response Protein-1 (EGR1) and GATA transcription factors in Figure 6c. EGR1 is involved in the control of cell physiology, which affects growth, division, and survivability. EGR1 is widely expressed in tissues and may be quickly activated by various environmental cues, including growth stimulants; shear stress, and reactive oxygen species. The study by Kumar et al., 2020 experimented that knocking out the EGR1 gene leads to prolonged survival of MTB in the host cell. Because MTB survival is strongly related to nitric oxide generation in murine macrophages, we examined nitrite production in infected macrophages [26]. GATA protein family members operate as lineage-specific transcription factors for a range of hematopoietic cell type systems. GATA2 regulates the transcription of genes involved in hematopoietic and endocrine cell lineage formation and proliferation. Mutation in the GATA2 gene leads to the development syndrome and is prone to mycobacterial infections [27].

The transcription factor analysis of the-miR-1234-3p had shown central clustering of expression levels with NRF2 (NEF2) and GATA1 [Figure 6d]. NEF2, a crucial modulator of the antioxidant defense system, was often found to protect against various lung illnesses. Human investigations have revealed that NEF2 plays an essential role in protecting against the oxidative damage caused by vigorous smokers. In healthy smokers, NEF2 was stimulated, and several antioxidant enzymes were boosted. The functionality of NEF2 in TB pathogenesis makes this transcription factor a unique entity and notable clinical gene for distinguishing TB patients from normal subjects [28].

We performed the analysis using gene expression of TB against antituberculosis medications, bedaquiline. The most significant contribution is the hub proteins that might be investigated as potential therapeutic targets and vaccination candidates. In this study, the hub genes identified are dnaA, dnaB, dnaE1, atpE, atpF, and atpB. During host infection, the increase of DNA gene expression directly contributes to mutation leads to DNA damage induced by immune-mediated reactive oxygen and nitrogen intermediates. Especially, ATP genes may play an essential role in the emergence of mutants that are better suited to surviving during infection and drug resistance evolution in this organism [29]. The bacteria tend to reduce cellular ATP consumption while increasing the capacity of ATP-generating pathways, which adds to bacterial survival in the face of antibiotic stress. In Mycobacteria, ATP synthase is required for growth, and the drug Bedaquiline effectively inhibits the operation of this critical metabolic enzyme. As a result, inhibitors of respiratory ATP production may be beneficial for eradicating this difficult-to-kill MTB subpopulation in human host cells [30].

4. CONCLUSION

This study identified the DEGs for the dataset GSE43749 and their associated biological processes, interactions, and pathways among healthy controls and active TB groups. The four miRNAs (hsa-miR-574-3p, hsa-miR-4787-5p, hsa-miR-601, and hsa-miR-1234) were determined to have a higher expression when susceptible to TB infection and primarily improve immunological responses. JUND, EGR1, NEF2, and GATA1 transcription factors were explored as a potential target gene. The observed hub genes seemed to have the most significant impact on ATP generation, DNA replication, immunological activity, and oxidative damage. Taken as a whole, the scientific community’s efforts are expected to usher in a new age of diagnostic tests based on hub genes – miRNAs and TF can be offered as prospective biomarkers. These diagnostic techniques must not only meet quality standards for specificity and sensitivity, but they must also be biologically relevant in the pathophysiology of TB.

5. AUTHORS’ CONTRIBUTIONS

V. Anusuya and P. Madhan carried out the experiments and prepared the original manuscript. B. Nivetha helped in editing and drafting the
manuscript. K Santhiya: Supervision, conceptualization, methodology, final review and approval.

6. FUNDING
There is no funding to report.

7. CONFLICTS OF INTEREST
The authors report no financial or any other conflicts of interest in this work.

8. ETHICAL APPROVALS
This study does not involve experiments on animals or human subjects.

9. DATA AVAILABILITY
All data generated and analyzed are included within this research article.

10. PUBLISHER’S NOTE
This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

REFERENCES
